SUFFICIENT CONDITION FOR TAP-LENGTH GRADIENT ADAPTION OF LMS ALGORITHM

Yuantao Gu, Kun Tang, and Huijuan Cui

State Key Lab on Microwave & Digital Communications,
Dept. of Electronics Engineering, Tsinghua University,
Beijing, 100084, CHINA. Email:gyt@mail.tsinghua.edu.cn

ABSTRACT

Besides the traditional accesses to accelerate the convergence of LMS algorithm, such as step-size control and input signal decorrelation, tap-length control is an emerging technique and attracts more and more attentions. Many tap-length control schemes are proposed and most of them are based on gradient search method. In this paper, the sufficient condition for tap-length gradient adaption is obtained based on the assumption of white Gaussian input. The analysis reveals that two requirements should be satisfied when use gradient method to search for optimum tap-length. One is that the unknown impulse response has a decay envelope, while the other requires that the number of the tap difference should be selected carefully.

1. INTRODUCTION

The well-known Least Mean Square (LMS) algorithm has been widely used in a variety of adaptive filtering applications such as echo cancellation, channel equalization, flow control, and multi-user detection, owing to the computational simplicity, numerical stability and ease of implementation [1]. However, low convergence rate limits its further popularity. Consequently, many variants, especially utilizing step-size control and input decorrelation, are proposed to accelerate the convergence rate of LMS algorithm.

Besides the two accesses mentioned above, tap-length control is another effective method to improve LMS, whose performance is significantly influenced by the taps number. To be specific, redundant taps would slow down convergence, accumulate more stochastic gradient noise, and aggravate computational cost. On the other hand, taps number must be large enough to reach acceptable identification accuracy. Therefore, there must exist an optimum tap-length that results in the fastest convergence rate.

How to select a fixed optimum tap-length had been studied very early [2], whereas the theoretical analysis of the influence of tap-length adaption on the convergence rate did not appear until 1990s [3, 4]. There are some control schemes [3, 5, 6, 7, 8], which aim to accelerate the initial convergence using a monotonically increasing tap-length sequence. Recently, some practical algorithms [9, 10, 11] that adapt tap-length in both directions are proposed.

This work was supported by the National Natural Science Foundation of China (60272020.) in part by the National Fundamental Research Program (G1998030406.)

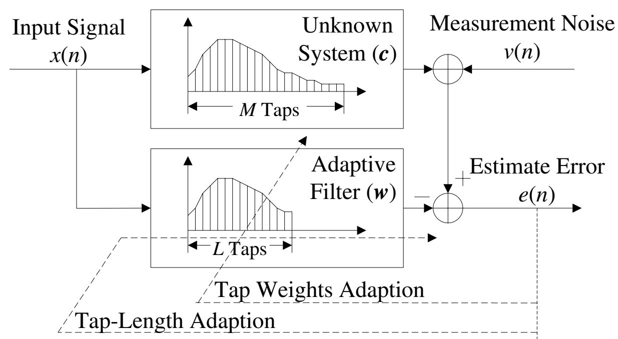

Fig. 1. System Identification with Variable Tap-Length Adaptive Filter

Most of the variable tap-length algorithms [9, 10, 11] are based on gradient method, which iteratively adapts the tap-length along the direction where cost function goes down. However, the condition for applying gradient method to search for optimum tap-length is not investigated in the available literatures.

In this paper, the convergence behavior of Mean Squared Estimate Error (MSEE) is drawn based on the assumption of white Gaussian input. Then the cost function of tap-length is defined as MSEE after one adaption. Consequently, the sufficient condition for tap-length gradient adaption is that the K-decimated cost function has unique minimum or monotonically varies, where $K \geq 1$ is the number of tap difference. By partitioning the unknown impulse response into blocks with same length, K, we prove that, if and only if, the energy of the partitioned response monotonically decreases, above condition can be satisfied. Therefore, decay envelope of impulse response and carefully selected K are the sufficient condition for tap-length gradient adaption.

2. MSEE CONVERGENCE WITH TAP-LENGTH ADAPTION

The work is presented using the set-up of system identification, see fig.1, where adaptive filter, $\mathbf{w}_L(n)$, is used to identify an unknown system, $\mathbf{c}_M$, with the constraint of minimizing a cost function of estimate error, $\xi(n) = \mathrm{E}\{e^2(n)\}$. In most literatures, the taps number of adaptive filter is considered exactly identical to that of unknown system, i.e., $D = |L - M| = 0$. The situation of truncation error, $L < M$, is studied in [12]. Generally, in the instance of $D \neq 0$, the optimum weights, $\mathbf{w}_L^*$, and minimum MSEE, $\xi^*(L)$, are, respectively,

$$\mathbf{w}_L^* = \begin{cases} \bar{\mathbf{c}}_L + \mathbf{R}_L^{-1}\mathbf{R}_{L,D}\underline{\mathbf{c}}_D, & L < M; \\ \left[\mathbf{c}_M^{\mathrm{T}}, \mathbf{0}_D^{\mathrm{T}}\right]^{\mathrm{T}}, & \text{otherwise.} \end{cases} \tag{1}$$

and

$$\xi^*(L) = \begin{cases} \sigma_v^2 + \underline{\mathbf{c}}_D^{\mathrm{T}}\left(\mathbf{R}_D - \mathbf{R}_{D,L}\mathbf{R}_L^{-1}\mathbf{R}_{L,D}\right)\underline{\mathbf{c}}_D, \\ \qquad\qquad\qquad\qquad L < M; \\ \sigma_v^2, \qquad\qquad\qquad \text{otherwise.} \end{cases} \tag{2}$$

where

$$\mathbf{c}_M = \begin{bmatrix} \bar{\mathbf{c}}_L \\ \underline{\mathbf{c}}_D \end{bmatrix}, \mathbf{R}_M = \begin{bmatrix} \mathbf{R}_L & \mathbf{R}_{L,D} \\ \mathbf{R}_{D,L} & \mathbf{R}_D \end{bmatrix}, \quad L < M,$$

denote, respectively, the partitioned matrix (vector) of unknown system response and input autocorrelation matrix, $\mathbf{R}_M = \mathrm{E}\{\mathbf{x}_M(n)\mathbf{x}_M^{\mathrm{T}}(n)\}$, and σ_v^2 denotes the variance of measurement noise. It can be readily recognized from (1) and (2) that the truncation error may reduce identification accuracy and yield biased result, unless the input signal is uncorrelated, i.e. $\mathbf{R}_M = \sigma_x^2\mathbf{I}_M$, where σ_x^2 denotes the input variance. On the contrary, the redundant taps produce an unbiasd result and does not bring any excess error to the minimum MSEE.

In LMS algorithm, $\mathbf{w}_L(n)$ is adapted iteratively along the stochastic gradient direction,

$$\mathbf{w}_L(n+1) = \mathbf{w}_L(n) + \mu e(n)\mathbf{x}_L(n), \tag{3}$$

where μ denotes adaption step-size and estimate error

$$e(n) = \mathbf{x}_M^{\mathrm{T}}(n)\mathbf{c}_M + v(n) - \mathbf{x}_L^{\mathrm{T}}(n)\mathbf{w}_L(n).$$

To arbitrary input signal, the convergence behavior of LMS is extremely sophisticated [1], even if the taps number is identical to that of unknown system. In this paper, the input is restricted to white Gaussian signal, which means the identification is unbiased in the scenarios of truncation error. Based on the well-known independence theory, the convergence of Mean Squared Weights Error (MSWE), $\epsilon(L,n) = \mathrm{E}\{\|\mathbf{w}_L - \mathbf{c}_M\|^2\}$ (with D zeros padded to the shorter vector), is eagerly derived [8],

$$\epsilon(L, n+1) = \beta\epsilon(L,n) + (\eta - \beta)t(L) + \gamma, \tag{4}$$

where β, η, and γ are parameters, and $t(L)$ denotes the truncation error,

$$\beta = 1 - 2\mu\sigma_x^2 + (L+2)\mu^2\sigma_x^4, \qquad \gamma = L\mu^2\sigma_x^2\sigma_v^2,$$

$$\eta = 1 + L\mu^2\sigma_x^4, \qquad t(L) = \begin{cases} \|\underline{\mathbf{c}}_D\|^2, & L < M; \\ 0, & \text{otherwise}. \end{cases}$$

Using the relation between MSEE and MSWE, $\xi(L,n) = \sigma_v^2 + \sigma_x^2\epsilon(L,n)$, MSEE converges as

$$\xi(L, n+1) = \beta\xi(L,n) + (\eta - \beta)\left(\sigma_v^2 + \sigma_x^2 t(L)\right). \tag{5}$$

The influence of tap-length on MSEE convergence is described by (5), from which the properties of optimum tap-length will be studied in the following sections.

3. SUFFICIENT CONDITION FOR TAP-LENGTH GRADIENT ADAPTION

3.1. Tap-Length Gradient Adaption

Many tap-length control schemes are based on gradient search method, which estimates MSEEs of two adaptive filters with adjacent tap-lengths, i.e., $\tilde{\xi}(L,n)$ and $\tilde{\xi}(L-K,n)$, and measures its difference,

$$\nabla\tilde{\xi}(L,n) = \frac{\tilde{\xi}(L,n) - \tilde{\xi}(L-K,n)}{K}, \tag{6}$$

where $K \geq 1$ is called *the number of the tap difference*. Subsequently, the tap-length is adapted along the negative difference direction,

$$L \Leftarrow L - \operatorname{sign}\{\nabla\tilde{\xi}(L,n)\}K. \tag{7}$$

According to (6) and (7), L adapts to (tracks) the minimum value of the sampled $\xi(L,n)$, on the condition that the sampled $\xi(L,n)$ has unique minimum or monotonically varies, where K is the sampling interval. However, in the available literatures there is no answer to the question – "under what circumstances, tap-length gradient adaption method can be applied?" – which seems more important than the adaption algorithm itself. Because dispensable tap-length adaption may computational expensive or lead to instability problems [13].

3.2. Optimum Tap-Length

The optimum tap-length, L^*, is defined as the one that minimizing MSEE after one iteration,

$$L^* = \arg\min_L \{\xi(L,n) | \xi(L', n-1)\}. \tag{8}$$

Please notice that L^* is time variant and independent to the tap-length before nth iteration, i.e., L'. According to the adaption rule of gradient search method, intuitively, L^* is the goal to which controlled tap-length iteratively tracks.

3.3. Relative Step-size

From (5), the behavior of LMS is determined by tap-length as well as step-size. To achieve the best convergence performance, combining tap-length control and step-size control is helpful. Consequently, a relative step-size is introduced and will be fixed during the adaption of tap-length [8],

$$\mu' = (L+2)\mu\,. \tag{9}$$

Definition (9) comes partly from the maximum step-size bound [8], $\max\{\mu\} = \frac{2}{(L+2)\sigma_x^2}$, and optimum initial step-size [4, 6], $\mu^* = \frac{1}{(L+2)\sigma_x^2}$, partly from Normalized LMS algorithm, in which step-size can be recognized as normalized to the energy of input data in the filter buffer. Actually, adjusting tap-length with relative step-size fixed is used in many variable tap-length algorithms [3, 4, 7, 8], though in some of them this is not explicitly declared.

3.4. Condition for Tap-Length Gradient Adaption

According to (8), L^* can be solved by set the difference of $\xi(L,n)$ (with respect to L) zero, with relative step-size fixed. Interval K is introduced to measure the difference. The unknown system response $\mathbf{c}_M$ can be also partitioned into m blocks with same length K,

$$\mathbf{c}_M^{\mathrm{T}} = \left[{\mathbf{c}_K^{(0)}}^{\mathrm{T}}, {\mathbf{c}_K^{(1)}}^{\mathrm{T}}, \cdots, {\mathbf{c}_K^{(m-1)}}^{\mathrm{T}} \right], \tag{10}$$

where $M = mK$. After some calculus (in appendix), the following two propositions are proved equivalent.

- *Proposition A: the energy of the partitioned unknown impulse response monotonically decreases, i.e.,*

  $$\|\mathbf{c}_K^{(i-1)}\|^2 \geq \|\mathbf{c}_K^{(i)}\|^2, \qquad \forall\, 0 < i < m.$$

- *Proposition B: the extremum of $\xi(L,n)$, if exists, is unique minimum.*

Recall section 3.1, proposition B is the sufficient condition for applying gradient search method, while proposition A demands the unknown response has decay envelope, as well as K is carefully selected. Consequently, the following theorem is derived.

Theorem: the sufficient condition for tap-length gradient adaption is that the unknown system response has a decay envelope, at the same time the number of the tap difference is carefully selected.

The proposed theorem firstly reveals that the unknown impulse response with decay envelope is required for applying gradient search method. As a matter of fact, such system is rather common in the physical world, i.e., acoustic echo path, and the transmission of wave. Therefore, gradient tap-length adaption algorithm can be widely used in many adaptive filtering applications. Furthermore, in the application of channel equalization or inverse identification, the unknown response is, generally, unimodal and decay in both directions. Correspondingly, the equalizer is centered and adds or removes taps in both side, where the theorem can be readily extended to. On the other hand, the sample applications presented in literatures are equalizers (or inverse identification) [3, 4, 5, 9], or with decay envelope [8, 10, 11], which confirm the proposed theorem.

The second problem disclosed by the proposed theorem is that the number of the tap difference is an important factor to tap-length gradient adaption algorithms. If a response with decay envelope is partitioned into two blocks, the condition of monotonically decreasing, i.e., $\|\mathbf{c}_K^{(0)}\|^2 \geq \|\mathbf{c}_K^{(1)}\|^2$, is certainly held. However, with the reduction of K, the condition may be unsatisfied because of the arbitrary response shape. Extremely, $K = 1$ and the response is M-partitioned, the condition definitely would not be satisfied in physical world. On the other hand, system performance will also deteriorate if K is too large, which results in a large deviation of the obtained tap-length from the optimal value. Therefore, the number of the tap difference should be carefully select and is dependent on application.

4. CONCLUSION

The energy of the partitioned unknown system response monotonically decreases is proved to be the sufficient condition for tap-length adaption applying gradient search method. The analysis results indicate that tap-length control algorithm can be applied a variety of applications with decay response envelope, whereas the number of the tap difference must be carefully selected.

5. APPENDIX: PROOF OF THE EQUIVALENCE BETWEEN TOW PROPOSITIONS

Firstly, MSEE is reproduced to a function of tap-length and relative step-size by utilizing (9) in (5),

$$\begin{aligned} \xi(L,n) &= \left(1 - \tfrac{2\mu'\sigma_x^2}{L+2} + \tfrac{\mu'^2\sigma_x^4}{L+2}\right)\xi(L', n-1) \\ &\quad + \left(\tfrac{2\mu'\sigma_x^2}{L+2} + \tfrac{2\mu'^2\sigma_x^4}{(L+2)^2}\right)\left(\sigma_v^2 + \sigma_x^2 t(L)\right). \end{aligned} \tag{11}$$

Notice that variable tap-length L is independent to the current MSEE, $\xi(L', n-1)$, and previous tap-length, L'. Then with μ' fixed, the first order difference of (11) with respect to L is got

$$\begin{aligned}\nabla\xi(L,n) &= \tfrac{\xi(L,n)-\xi(L-K,n)}{K} \\ &= 2\mu'\sigma_x^4\left(\tfrac{J}{L(L-K)}+\tfrac{t(L)}{LK}-\tfrac{t(L-K)}{(L-K)K}\right),\end{aligned} \tag{12}$$

where K is the number of the tap difference and

$$J=\frac{(2-\mu'\sigma_x^2)\xi(L',n-1)-2\sigma_v^2}{2\sigma_x^2}\,.$$

In (12) and the following calculus, L is assumed large enough, $L \gg 4\max\{\mu'\}\sigma_x^2 = 8$, which is generally held for applications using tap-length adaption. Some items in (12) with L^2 in denominator are neglected because they are so small with respect to the others. Similarly, $L+2$ is approximated to L.

The constraint of extreme, L_{ext}, is obtained by set (12) zero,

$$KJ+(L-K)t(L)+Lt(L-k)=0\,. \tag{13}$$

The second order difference must be evaluated to determine if L_{ext} is minimum or maximum,

$$\begin{aligned}\nabla^2\xi(L,n) &= \tfrac{\nabla\xi(L,n)-\nabla\xi(L-K,n)}{K} \\ &= 2\mu'\sigma_x^4\left(\tfrac{-2J}{L(L-K)(L-2K)}+\tfrac{t(L)}{LK^2}\right. \\ &\quad \left.-\tfrac{2t(L-K)}{(L-K)K^2}+\tfrac{t(L-2K)}{(K-2K)K^2}\right).\end{aligned} \tag{14}$$

Then utilizing (13) in (14), the second order difference at L_{ext} is achieved,

$$\nabla^2\xi(L,n)\big|_{L_{\text{ext}}} = 2\mu'\sigma_x^4\tfrac{t(L)-2t(L-K)+t(L-2K)}{(L-2K)K^2}\,. \tag{15}$$

If the unknown system response is partitioned as (10), the truncation error can be rewritten to

$$t(L)=\begin{cases}\sum_{i=\frac{L}{K}}^{m-1}\|\mathbf{c}_K^{(i)}\|^2, & L<M\,;\\ 0, & \text{otherwise}\,.\end{cases} \tag{16}$$

Consequently, using (16) in (15), one gets

$$\nabla^2\xi(L,n)\big|_{L_{\text{ext}}} = \frac{2\mu'\sigma_x^4\left(\|\mathbf{c}_K^{(\frac{L}{K}-2)}\|^2-\|\mathbf{c}_K^{(\frac{L}{K}-1)}\|^2\right)}{(L-2K)K^2}\,, \tag{17}$$

It can be readily obtained that $\xi(L,n)\big|_{L_{\text{ext}}}$ will be a minimum, if and only if, $\|\mathbf{c}_K^{(i)}\|^2$ monotonically decreases. Because it is impossible that there has no maximum, but more than one minimum. Therefore, the extreme, if exists, must be unique minimum. The proof to the mentioned equivalence is closed.

6. REFERENCES

[1] S. Haykin, *Adaptive Filter Theory, 4th ed*, Beijing: Publishing House of Electronics Industry, 2002.

[2] W. S. Hodgkiss, "Selecting the length of an adaptive transversal filter," in *ICASSP '78*, Apr. 1978, vol. 3, pp. 96–99.

[3] Z. Pritzker and A. Feuer, "Variable length stochastic gradient algorithm," *IEEE Trans. Signal Processing*, vol. 39, pp. 997–1001, 1991.

[4] K. Wesolowski, C. M. Zhao, and W. Rupprecht, "Adaptive lms transversal filters with controlled length," *IEE PROCEEDINGS-F*, vol. 139, pp. 233–238, June 1992.

[5] Y. K. Won, R.-H. Park, J.-H. Park, and B.-U. Lee, "Variable lms algorithms using the time constant concept," *IEEE Trans. Consumer Electron.*, vol. 40, pp. 655–661, Aug. 1994.

[6] V. H. Nascimento, "Improving the initial convergence of adaptive filters: variable-length lms algorithms," in *International Conference on Digital Signal Processing*, 2002, vol. 2, pp. 667–670.

[7] R. C. Bilcu, P. Kuosmanen, and K. Egiazarian, "A new variable length lms algorithm: theoretical analysis and implementations," in *International Conference on Electronics, Circuits and Systems*, 2002, vol. 3, pp. 1031–1034.

[8] Y. Gu, K. Tang, H. Cui, and W. Du, "Convergence analysis of a deficient-length lms filter and optimal-length sequence to model exponential decay impulse response," *IEEE Signal Processing Lett.*, vol. 10, pp. 4–7, Jan. 2003.

[9] F. Riera-Palou, J. M. Noras, and D. G. M. Cruickshank, "Linear equalisers with dynamic and automatic length selection," *Electronics Letters*, vol. 37, pp. 1553–1554, Dec. 2001.

[10] Y. Gu, K. Tang, and H. Cui, "Lms algorithm with gradient descent filter length," *IEEE Signal Processing Lett.*, #SPL4501, in press, will be published in 2004.

[11] Y. Gong and C. F. N. Cowan, "Structure adaptation of linear adaptive filters," *submitted to IEE Proceedings - Vision, Image and Signal Processing*.

[12] P. Eykhoff, *System Identification: Parameter and State Estimation*, London: New York, Wiley-Interscience, 1974.

[13] R. D. Poltmann, "Stochastic gradient algorithm for system identification using adaptive fir-filters with too low number of coefficients," *IEEE Trans. Circuits Syst.*, vol. 35, pp. 247–250, Feb. 1988.